Wireless Sensing Used for ASL Detection: A Survey

Leena Bacha
School of Engineering
University of Virginia
Email: snk6kh@virginia.edu

Sarah Boyce
School of Engineering
University of Virginia
Email: abp6cm@virginia.edu

Ganesh Nanduru
School of Engineering
University of Virginia
Email: bae9wk@virginia.edu

***Abstract*—Recent developments in wireless sensing have demonstrated the potential to bridge communication between the deaf and hearing communities through automated American Sign Language (ASL) recognition. We analyze current distinct wireless sensing systems using Wi-Fi, RFID, acoustic waves, and millimeter-wave technologies, evaluating their recognition capabilities, technical limitations, and ease of implementation. We draw unique insights by comparing results across the different wireless sensing systems to identify the strengths and drawbacks of each general system. This survey also emphasizes the practicality of each design by reporting each study's cost and complexity. Our main discoveries are there is no single ASL recognition system that is significantly better in all criteria than the rest, and each system can benefit in accuracy and reliability greatly from incorporating state-of-the-art AI to interpret received signals.**

I. Motivation

ASL (or American sign language) is a key piece in bringing inclusivity to the deaf community. By extending this to the use of different wireless sensing techniques, we can improve communication between those who do and don't know ASL. Analyzing this concept from a wireless sensing approach offers a contactless, non-invasive method that can detect subtle gestures, imperative for ASL.

This problem is important because as we advance in real time translation from spoken language to spoken language, it is important we keep up with ASL as well as spoken language to further protect the inclusion of the deaf community. This is not only useful for those who use ASL, but also for those who want to communicate with the hearing impaired. This can also have an extension across different types of sign language and thus different spoken languages as well.

II. Introduction

ASL is not just a mode of communication for the hard of hearing community but also a bridge for fostering inclusivity in society. Recent advancements in real-time language translation have predominantly focused on spoken languages, leaving a gap in similar technologies for sign language. Addressing this gap is crucial to ensuring equitable communication tools that serve both ASL users and those wishing to engage with them. Wireless sensing technologies offer a unique solution, providing contactless, non-invasive methods capable of detecting the subtle gestures essential for ASL. Utilizing wireless sensing technologies, a person's privacy is retained as the user does not need to be continuously recorded as only reflections back to the sensor are measured[11]. Unlike camera based methods, wireless sensing is also less affected by environmental factors such as low lighting, ensuring consistent performance in diverse settings [11][14]. Furthermore, wireless systems can operate in real-time and at greater distances, offering enhanced flexibility for practical applications in everyday environments.

This paper examines six wireless sensing systems spanning technologies like Wi-Fi, RFID, acoustic waves, and mmWave comparing their capabilities, limitations, and potential for advancing ASL recognition. By offering a detailed comparative analysis, this review aims to identify the strengths and trade-offs of each approach, guiding future innovations in inclusive communication tools.

A. Metrics

ASL is a specific type of human detection. It does not focus on human vitals or the movements of an entire body, it specifically focuses on a small portion of the human body, the hands. In order to translate from ASL, minute gestures and subtle differences in those gestures are imperative to proper detection. When considering different sensing techniques, we look at three metrics in order to classify its effectiveness as an ASL detection system.

1) **Classification Accuracy**: This is a percentage on how well a system is able to detect and therefore classify hand gestures or positions to the correct letter, word, or phrase it is trying to achieve. As we are looking for a system that interprets ASL as correctly as possible. We will have to come up with a method to compare accuracies between different experimental setups, as some papers achieve higher accuracies by including less classes in their trials, which demonstrates accurate classification within the context of the experiment but lacks generalizability.
2) **Cost**: Another key factor in evaluation is cost. A good ASL detection system should be made accessible through cheap implementation. Our estimation of cost depends largely on the thoroughness of documentation done by each researcher, as without exact prices we are left to guess which devices were bought to perform the experiments and at which price.

3) **Reliability**: Finally, we are looking for systems with proven robustness in diverse environments, with variations in weather conditions, hand/arm size and shape, and cultural backgrounds of the ASL communicators.

III. WiFi Sensing

Wi-Fi Sensing works by detecting and classifying disruptions in the Wi-Fi spectrum caused by movement over time. The Wi-Fi clients emit electromagnetic waves in the form of standard communication data frames, which are used for sensing. These signals are transmitted back to the AP at a configurable sounding rate, which can be adjusted for faster or slower measurements.

A. WiFinger

WiFinger leverages WiFi signals to recognize fine-grained finger gestures, providing a contactless and non-invasive approach to human-computer interaction. By analyzing Channel State Information (CSI) signal changes caused by 9 different finger movements, WiFinger identifies unique gesture patterns. The system uses a wireless access point and a laptop to capture these variations. Through noise removal, filtering, and advanced algorithms like dynamic time warping (DTW) and k-Nearest Neighbors (kNN), WiFinger processes and classifies these patterns into gestures. Although WiFinger cannot detect the full scope of ASL, as it only classifies hand symbols representing numbers, it is a step in the right direction.

The system's cost-effectiveness is a standout feature, as it relies on commercially available WiFi devices rather than requiring specialized hardware like software-defined radios or wearable sensors. By simply modifying existing WiFi drivers, WiFinger offers a scalable and affordable solution for gesture recognition, which can be widely adopted without significant financial barriers.

In terms of reliability, WiFinger demonstrates resilience in controlled environments, effectively filtering out noise and adapting to minor disturbances. However, its performance can be impacted by dynamic changes, such as moving objects or additional people near the user, which can interfere with the CSI signal patterns. While these limitations highlight areas for improvement, WiFinger's robust preprocessing and feature extraction techniques allow it to maintain reliable operation under moderately stable conditions.

WiFinger achieves an average recognition rate of 90.4% for the nine individual hand gestures. To test its accuracy in sequences of gestures, 20 sequences were created, each containing 18 gestures. The test subjects then performed 5 of the 20 sequences at random. This resulted in a 82.67% classification accuracy. [5]

B. Sign Language Gesture Recognition using Doppler Radar and Deep learning

In this system, a microwave X-band Doppler radar is used to capture ten emergency specific hand gestures by detecting micro-Doppler signatures or unique frequency variations caused by hand movements. The radar transceiver transmits a signal that reflects off the user's hands, capturing motion variations as raw data. This data is processed using joint time-frequency analysis to generate spectrograms that visually represent the gesture patterns. These spectrograms are then classified using deep learning algorithms.

Initially, a Deep Convolutional Neural Network (DCNN) is employed to analyze the spectrograms, achieving high classification accuracy. To further enhance performance, a pre-trained VGG-16 transfer learning model is applied, improving accuracy by leveraging its ability to identify subtle patterns in images. This system effectively combines hardware-based radar sensing with advanced machine learning techniques to provide a non-invasive and accurate method for recognizing ASL gestures. This is more closely a representation of ASL as this system is detecting more than just hand positions, but rather entire movements.

This study employs accessible and relatively affordable microwave X-band Doppler radar hardware paired with Matlab and LabVIEW software. By using commercially available equipment, the approach avoids the need for specialized or costly alternatives, making it a practical choice for widespread adoption.

The performance can vary depending on factors such as user proximity to the radar and environmental noise. The preprocessing steps and use of advanced algorithms, like the Deep Convolutional Neural Network (DCNN), enhance robustness but leave room for optimization in more dynamic environments or with more diverse user profiles.

In terms of accuracy, the results are impressive. The DCNN algorithm achieves an 87.5% validation accuracy for classifying 10 distinct ASL gestures, while applying the VGG-16 transfer learning model improves this to 95%. This high accuracy level highlights the potential of Doppler radar combined with deep learning for real-world applications. [2]

C. Other WiFi Sensing Methods

To supplement our research in WiFi sensing methods, we analyze the findings of another literature review[7]. We found overviews of two additional methods to be prominent:

1. WiCatch uses channel state information (CSI) to construct a virtual antenna array to track hand movements, achieving 95% recognition accuracy on 9 two-hand gestures. WiCatch does not require any wearable sensors or additional equipment beyond a wireless link card, which can be purchased for under $50 (e.g. Intel 5300 NIC) [12].

2. WiSee introduces an alternative gesture recognition system using Doppler-based detection. By constructing a profile out of Doppler shifts extracted from wireless signals, researchers were able to match Doppler shift patterns to hand gestures. Using this technique, WiSee achieved 94% accuracy on 9 gestures. However, WiSee researchers relied on USRP-N210 radios to achieve these results, and these radios can cost upwards of $3,500.00 each [10].

IV. RFID

Radio frequency identification (RFID) is a wireless detection scheme that uses tags and readers to monitor an environment. RFID tags are microchips that can be active (battery powered) or passive. These tags are light, small, and can be worn or placed on objects meant to be tracked. RFID readers will read passive tags by transmitting energy to them and decoding the response given by the chip. This is generally close-range, with typical passive RFID systems working between 10 cm - 1 m of range. Active RFID systems use powered tags that transmit a signal at a much further range, however passive tags are significantly cheaper to purchase [4].

A. *RF-Sign*

RF-sign is an RFID-based gesture recognition system that aims to improve detection accuracy by capturing granular finger movements as opposed to coarser hand movements. Researchers designed a glove with passive RFID tags on each finger and attempted to track the phase of each tag to reconstruct the user's finger movements. However, researchers encountered two major roadblocks in implementing this design:

1. Continuous gestures cannot be segmented easily using empirical thresholds. As the ASL user's wrist orientation changes during signing, the spatial readings of the RFID tags change in an uneven manner, making it difficult to interpret the tracked movements of the fingers. To address this, researchers standardized the wrist orientation by adding an additional tag on the wrist of the glove. This tag acts as a reference to the finger tags, making it much easier to interpret changes in finger position as the subject moves.

2. Finger movements across different axes relative to the reader antennas register differently. For example, the researchers discovered that finger on the axis perpendicular to to the antenna pair plane exhibit different degrees of change to the phase profile when compared to the same movements executed on the axis parallel to the antenna pair plane. To resolve this, the researchers designed a phase matrix reconstruction algorithm to be position-invariant by first confirming the hand's rotation angles, then using the hand's orientation as a reference to reconstruct the finger movements.

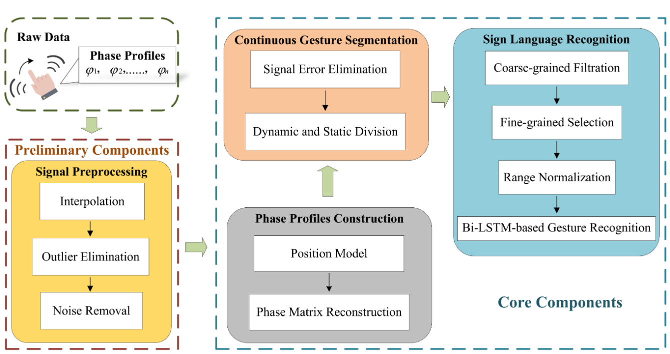

Fig. 1. RF-Sign system architecture [14]

Fig. 1 shows the system architecture of RF-sign, divided into four major components. Researchers outline phase profiles construction as a key contribution of this paper, as both major problems encountered by the researchers were solved by recalibrating phase profiles using knowledge of the signals' context, whether it was hand rotation or the position of the RFID tag on the user's wrist.

RF-sign achieves a 92.81% recognition accuracy on the alphanumeric hand signals with an average angle error of $4.3°$. It is able to recognize certain signs such as Z, X, and 9 with 100% accuracy, and can recognize 21 out of 26 letters with 91% accuracy. RF-sign even tracks gestures accurately with obstacles placed between the antennas and glove. However, the system experiences a drop in accuracy once the glove is more than 150cm away from the antennas. This result suggests that a system like RF-sign would be most effective in cases where the reader and signer are positioned closely, which may be possible in controlled environments but not as feasible out in the open.

In terms of cost, RF-sign is one of the cheapest implementations in this study. Passive RFID tags generally cost under a dollar each, and the corresponding reader can be purchased for roughly $40.00. Compared to the other methods discussed in this survey, RF-sign is cheap and accurate at the cost of low range and limited use cases, scoring well in the classification accuracy and cost metrics but falling relatively short in reliability. [14]

V. Millimeter Wave Sensing

Millimeter-wave (mmWave) sensing takes advantage of high-frequency electromagnetic waves to enable precise detection, tracking, and imaging of objects in various environments. mmWave utilizes FMCW (Frequency Modulated Continuous Wave) radar signals to determine an object's range, angle, and velocity. This process occurs when a frame is sent by a transmitter and received by a receiver which computes the range using a range fast fourier transformation (FFT) and performs a Doppler estimation. A CFAR (Constant False Alarm Rate) is applied to the output of the Doppler FFT for object detection. From this, the direction of arrival is estimated resulting in the ability to calculate the azimuth and elevation values of the object. These results are useful in many scenarios and can be specifically applied in an ASL recognition system by understanding the movement of hand gestures through this estimation.

A. *Expressive ASL Recognition using Millimeter-wave Wireless Signals*

ExASL utilizes mmWave sensing to recognize both manual markers (hand gestures) and non-manual markers (head and torso movements). Developed by researchers at George Mason University in 2020, this system separates and identifies body parts to accurately detect and interpret both types of markers in ASL. Non-manual markers are crucial for understanding the full meaning of sentences in ASL, as phrases and sentences can be significantly augmented or altered through subtle head and torso movements. An example demonstrating this phenomenon is the sentence "I like apples". If a person were to sign this phrase but shake their head in a negative manner, then the sentence transforms into “I do not like apples” even though the gesture remains the same.

There are three main components to ExASL, mmWave point cloud generation, multi-distant clustering, and multi-view deep learning. During this process an mmWave transmits signals that reflect off the user’s body when signing. Based on information gathered from the gesture’s range, angle, and velocity, for multiple points on the user’s body, a 3D point cloud is created based on these points. A multi-distance clustering algorithm is run on the point cloud to separate and define corresponding body parts (left and right hand, head, and torso) based on predefined Kinect templates. Finally, a multi-view CNN (Convolutional Neural Network) extracts the spatial features from each frame while a Long Short-term Memory (LTSM) analyzes the temporal evaluation of these features for both manual and non-manual marker recognition. Non-manual markers are recognized through specific head and torso movements. These include torso shifts forward for Yes/No questions and Wh- questions (accompanied by a head tilt), a side-to-side head shake for negation, a head nod for assertion, a torso shift away for verb inflection, and a torso shift right then left for spatial agreement (used to identify multiple subjects or objects within a single sentence). This entire process is necessary in order for ExASL to efficiently and accurately recognize manual and non-manual markers in ASL sentence recognition.

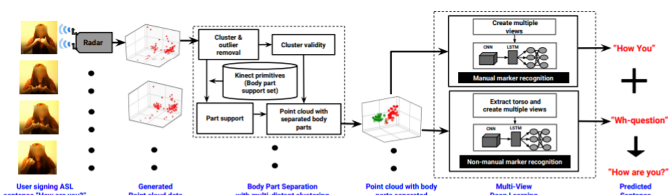

Fig. 2. Workflow of ExASL when performing sentence level ASL recognition [11]

In order to test the efficacy of ExASL, 5 participants trained and tested a dataset of 23 signs and 29 sentences made up of these signs on three models: Unclustered multi-view which does not separate body parts, Clustered multi-view which utilizes body part separation, and Clustered multi-view swapped which swaps left and right hands for every input doubling training sample size. Overall, Clustered multi-view swapped achieved the highest accuracy for both word-level recognition (92.5%) and sentence-level recognition (WER of 0.79%, SER of 1.25%). This model’s remarkable performance suggests that there are benefits to having body part separation and data augmentation in enhancing ASL recognition accuracy. All models exhibited difficulty differentiating between words with similar gestures, a challenge that persisted in sentence-level recognition. There was also a noticeable increase in word and sentence level error when cross subject evaluation occurred (utilizing a different participant than the model was trained on) indicating the need for greater participant diversity in training data. [11]

B. Other mmWave Sensing Methods

While researching a literature survey for gesture recognition, we encountered Soli, a high-resolution and low-power mmWave radar-based solution developed by Google in 2016. Soli achieved 92.1% accuracy on 4 hand gestures with minimal power consumption while running at over 10,000 frames per second [6].

VI. Acoustic Sensing

Acoustic sensing takes advantage of sound waves to detect and interpret gestures by analyzing the changes in wave propagation caused by hand and body movements. Inaudible acoustic sounds are emitted and the signals reflected back are analyzed to perform different functions. With acoustic sensing, hand gesture recognition, hand gesture tracking], localization, user authentication, keystroke snooping attacks, and environmental sounds are all able to be captured and performed [1].

A. HearASL

HearASL employs acoustic sensing through a smartphone system in order to recognize ASL words and sentences. Developed by researchers at the Beijing Institute of Technology and Temple University in 2022, this system analyzes inaudible sound waves emitted from a smartphone’s speaker and records the reflection of the signer’s hands through the smartphone’s microphone. To the researchers’ knowledge, it is the first smartphone-based ASL recognition system and is designed to be resistant to ambient noise.

The HearASL process comprises three main stages: data collection, CIR enhancement, and sign language recognition. In data collection, the transmitted signal, modulated with a Barker code for precise channel estimation, is sent out by the speaker, and the received signal is filtered to eliminate ambient noise. During CIR enhancement, the system estimates the channel impulse response (CIR), detailing how the signal reflects off the signer’s hands and returns to the microphone. The CIR is converted into an image where time, distance, and signal strength are represented on the x-axis, y-axis, and pixel brightness, respectively. Static reflections from stationary objects are removed using a 1st-order difference operation, and a Hampel filter reduces noise and outliers. Finally, sign language recognition employs deep learning models for word and sentence-level classification. A CNN with an attention mechanism is used for word-level recognition, while a CNN+ gate recurrent unit (GRU) + connectionist temporal classification (CTC) model handles sentences. The GRU layer helps capture temporal dependencies between signs, allowing the model to understand the context of the sentence while the CTC loss function allows the model to learn directly from unsegmented data, eliminating the need to manually label the start and end times of each word in a sentence. HearASL further improves the accuracy of sentence-level recognition by including specific labels for common transition movements that occur between words, ensuring that these movements are ignored during the recognition process.

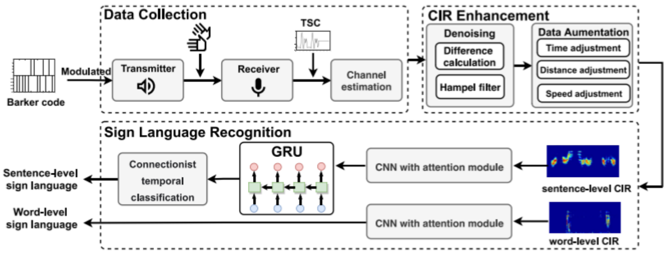

Fig. 3. Workflow of HearASL when performing ASL recognition [13]

HearASL demonstrated impressive results in recognizing and translating ASL. In user-dependent tests, the system achieved 97.2% accuracy for individual signs and a 0.9% WER for sentences. In user-independent tests, the accuracy remained high at 90.8% for one-handed signs and 90.6% for two-handed signs, with a WER of 3.4% in sentences. The system also underwent extensive robustness testing, including evaluations in various environments, across different distances, and at varying angles. Tests conducted in real-world settings such as apartments, corridors, sidewalks, and restaurants showed that accuracy only slightly decreased in noisy environments. The system's performance remained consistent across a range of distances (20-80 cm) and angles (0°-60°) between the signer and the smartphone. HearASL performs optimally at a distance of 40 cm between the signer's hands and the smartphone, with accuracy declining at greater distances due to weaker signal strength and increased challenges in channel estimation. Regarding angle variations, HearASL maintained highest accuracy at 0°, decreases as angle increases, but remains relatively in optimal ($> 87.5\%$) range at 40°. [13]

B. Other Acoustic Sensing Methods

Another literature review we read made reference to an acoustic-based intrusion detection system that operated on signals acquired by a distributed microphone network. While this system benefits from higher spatial resolution due to audio waves having smaller wavelengths than WiFi signals, a distributed microphone network solution for ASL recognition would require complex setup and would have a higher cost of implementation. Moreover, the acoustic detection system referenced in the survey was designed to detect intruder movements within a building, which is a task much less granular than the finger-level of detail required for ASL recognition [7] [15].

VII. Discussion

Having explored the diverse wireless sensing technologies utilized for ASL recognition, this section delves into a comparative analysis of these approaches, highlighting their respective strengths, limitations, and applicability across various scenarios. All the systems presented address the privacy and low light concerns associated with using cameras, as wireless sensing enables recognition without requiring the person to be visibly recorded.

A. Classification Accuracy

Accuracy is a critical factor in evaluating the ability of wireless sensing technologies to perform ASL recognition. High accuracy is essential for ensuring reliable communication and practical usability, particularly in real-world scenarios. The accuracy of each model was based on its highest-performing configuration, and the results are summarized in Table VII-C. Although all the papers include some form of accuracy measurement, the metrics are based on different aspects of hand signaling, such as hand positions, alphanumeric gestures, words, or sentences, which vary across the studies. HearASL demonstrated the highest accuracy overall, achieving 97.2% accuracy for word-in-sentence recognition in private settings and optimal positioning relative to the phone sensor [13]. This is followed by SignFi, which achieved 94.81% accuracy [8] across an impressive 276 gestures. The Doppler based deep learning model reported an accuracy range of 87.5% to 95%, depending on whether transfer learning with VGG-16 was applied [2]. RF-Sign and ExASL both achieved roughly 93% accuracy, with RF-Sign focusing on alphanumeric recognition and ExASL performing word-level recognition across 23 words[14][11]. Finally, WiFinger, while having the lowest accuracy at 90.4%, still maintains a strong performance for ASL recognition[5]. These results indicate that while each system demonstrates high accuracy in its target domain, the differences in scope have become a critical factor in determining the most suitable option for specific applications. Systems like HearASL and SignFi are particularly notable for balancing both high accuracy and scope of recognition, showcasing their potential for practical deployment.

B. Cost

An important factor to consider when evaluating wireless sensing technologies for ASL recognition is the cost of implementation, which includes hardware requirements, setup complexity, and potential scalability for widespread deployment. The Doppler detector-based device utilizes a microwave HB100 radar detector, which costs roughly $7, alongside a DAQ device priced between $79 and $1,000, depending on specifications. Both SignFi and WiFinger leverage the existing infrastructure of commercial Wi-Fi devices, incurring little to no additional cost if a router is already owned. RF-Sign uses RFID tags, which are highly affordable (approximately $14 for 50 tags), paired with sensors costing $30–$50. ExASL employs a commercially available mmWave radar from Texas Instruments, priced between $20 and $35. Although ExASL has a relatively low hardware cost, its training cost is notable. Training the word-level model required approximately 14 hours for 23 words, while the sentence-level model took about two days to train on 29 sentences, as it needs to perform both body part separation and processing through a CNN+LSTM system [11]. This cost would increase significantly with the addition of more words and phrases to train a comprehensive ASL dataset. HearASL, on the other hand, implemented its system using the iPhone 12 Pro, but its approach should theoretically work on any modern smartphone with a microphone and speaker system, which makes it one of the most scalable and accessible solutions due to the widespread availability of smartphones. From this analysis, it becomes clear that while hardware costs for most systems are relatively low, training and scalability considerations introduce significant trade-offs. Systems like WiFinger and HearASL, which rely on existing devices, present the most cost-effective and scalable solutions. However, more specialized systems, including ExASL, offer unique capabilities at the expense of higher training complexity, highlighting the importance of aligning system choice with application-specific needs.

C. Reliability

Another factor in evaluating these systems is their reliance and robustness, as their effectiveness in real-world scenarios depends on consistent performance across diverse environments, users, and conditions. In controlled environments, WiFinger demonstrates a promising level of resilience, effectively filtering out noise and adapting to minor disturbances. Its preprocessing and feature extraction techniques enable reliable operation under moderately stable conditions[5]. However, its performance can degrade in the presence of dynamic changes, such as moving objects or additional people near the user, which can interfere with CSI signal patterns [5]. The robustness and limitations of the doppler-based system were not explicitly addressed in the study, but other sources suggest that it may be susceptible to interference from other moving objects and variations in user speed and distance, which could degrade its performance [7].

ExASL's separating body parts method reduces the impact of second-order reflections and noise, achieving high accuracy in recognizing signs and sentences. RF-Sign demonstrates robustness by utilizing position models and reference tag RSS trend analysis, ensuring accurate recognition despite variations in angles, distances, and movement speeds [14]. HearASL, which already uses inaudible acoustic signals that operate beyond the range of urban noises and human speech, has evaluated its robustness by testing performance under varying angles, distances, and levels of environmental noise. These tests revealed that HearASL can achieve high accuracy in word and sentence recognition in various settings, such as restaurants, and determined optimal distance and angle ranges [13]. Each of these methods highlights a unique approach to improving reliability and robustness in ASL recognition. However, continued development of systems that adapt to real-world variability and environmental challenges is essential to achieve truly reliable ASL recognition technologies.

VIII. Conclusion

The original goal of this literature review was to identify the best ASL recognition system utilizing wireless sensing available at present. However, the papers reviewed reveal a diverse landscape of sign language recognition systems, each with distinct strengths and limitations. While some systems focus on recognizing individual signs, such as alphanumeric symbols, others achieve complete word and sentence recognition. As a result, the choice of the most suitable system depends on specific application requirements, including the level of detail needed, cost constraints, desired robustness, and the target user population. Notable standouts from this review include SignFi, ExASL, and HearASL, all of which demonstrate high accuracy ($> 92\%$) in gesture, word, or sentence recognition. SignFi is particularly notable for its ability to learn 276 gestures, ExASL excels in recognizing sentences along with non-manual markers, and HearASL achieves the highest recognition accuracy with a larger dataset comprising 50 words and 30 sentences and performing robustness tests. Compared to an existing survey of wireless sensing techniques for gesture recognition [7], this literature review better represents the capabilities of acoustic sensing for ASL. In comparison, the gesture recognition survey we analyzed frames acoustic sensing as neither low-cost nor non-intrusive, which may not be accurate. Additionally, this paper provides more specific context for the results of other studies, reporting gesture type, number word/sentence classes, and individual system drawbacks to add more depth to the final accuracy reported by each experiment surveyed.

IX. Recommendations and Future Work

Based on the information gathered and analyzed, we have identified several key considerations for implementing an ASL recognition system using wireless sensing. These include creating a mixed word and finger position system, improving handling of user variability, expanding sign language datasets to encompass a broader range of signs, enhancing environmental robustness for real-world scenarios, and incorporating non-manual marker recognition. A common trend across existing systems is either word/sentence recognition or alphanumeric/gesture recognition, but none have effectively combined both. Since ASL involves both letter signing and word/sentence signing, it is crucial to address this gap. Another challenge identified in deep learning-based systems is the drop in accuracy when users who did not participate in training the model perform sign to the model, highlighting the need for more diverse user data to improve model robustness. This is exemplified in the HearASL study where including a higher percentage of training data resulted in having better performance [13]. Furthermore, increasing the number of words, sentences, and gestures in training datasets is vital, as current systems typically train on only 50 words or fewer which is far fewer than what is required for effective communication. Notably, sign language involves not only manual hand signs, but also non-manual markers, which are critical for accurate communication. Without the inclusion of these markers, even systems capable of recognizing a wide range of words and sentences will miss significant aspects of ASL. Lastly, for real-world applications where users may not understand ASL, ensuring environmental robustness is essential. The system must be able to filter out background noise and adapt to various environmental conditions to facilitate reliable communication in diverse settings.

TABLE I

*NOTE: THERE IS A DIFFERENCE HERE BETWEEN SEQUENCE AND SENTENCE. SIMILAR TO ENGLISH, A STRING OF WORDS IS NOT NECESSARILY A SENTENCE.

Device Name	WiFinger	SignFi	Doppler with Deep Learning	RF-sign	Expressive ASL	Hear-ASL
Year of Publication	2013	2018	2019	2023	2020	2022
Signal	WiFi	WiFi	WiFi	RFID	mmWaves	Acoustic
Device	WiFi Antenna	WiFi Antenna	Radar	RFID Tags	Radar	Smartphone
Hand Gesture Type	Position	Movement	Position	Position	Movement	Movement
No. Words	9	276	10	36	23	50
No. Sentences	20	N/A	N/A	N/A	29	30
Accuracy	90.4%	94.81%	87.5%-95%	92.81%	92.5%	97.2%

As we consider the potential impact of AI on ASL recognition, it is important to examine how recent advancements in deep learning could shape the future of sign language recognition, particularly in the context of recognizing entire ASL gestures rather than just individual hand positions or specific gestures. When looking further into deep learning for ASL, there are no significant changes compared to what we have seen so far. [3] is using deep learning to classify gesture movements (or words and phrases), which is an advancement, however it is not for ASL but rather Chinese sign language. The other research for ASL is classifying hand positions using deep learning [9], but as previously stated, hand positions is not representative of the language, rather just letters. Due to this classification technique, there has been advancements in accuracy, over 99%, which is an improvement in the techniques discussed here, but, we are looking for advancements in the entire language, not just gesture recognition. Due to this, we have not included recent research in this area, but encourage the reader to investigate on their own using these papers. [3] [9].

X. Individual Contributions

All members within the group were responsible for reading and encapsulating two research papers and analyzing their classification, cost, and reliance. We each summarize our individual contributions:

- Leena: I covered the millimeter wave and acoustic wireless sensing sections. This includes encapsulating the ExASL and HearASL papers as well as analyzing their accuracy, cost, and reliability. I also wrote the discussion between all the papers, introduction before metrics, conclusion, and provided the recommendations in the recommendations and future work sections.
- Sarah: I added the motivation section and the metrics to the introduction. In addition I wrote the forward to the WiFi sensing section as well as read, summarized and analyzed the first two WiFi papers. I created the analysis table and added the last paragraph to the recommendations and future work. In general, I organized the layout and format of this report.
- Ganesh: I analyzed RF-sign to investigate the capability of RFID-based detection systems in recognizing ASL. I also report a brief summary on RFID, and I summarized results from another literature review on state-of-the-art wireless sensing for gesture recognition throughout this paper. I also wrote the abstract of this paper.

References

[1] Yang Bai et al. "Acoustic-based sensing and applications: A survey". In: *Computer Networks* 181 (2020), p. 107447. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2020.107447. URL: https://www.sciencedirect.com/science/article/pii/S1389128620311282.

[2] Hovannes Kulhandjian et al. "Sign Language Gesture Recognition Using Doppler Radar and Deep Learning". In: *2019 IEEE Globecom Workshops (GC Wkshps)*. 2019, pp. 1–6. DOI: 10.1109/GCWkshps45667.2019.9024607.

[3] Yi-Chen Lai, Pin-Yu Huang, and Tzyy-Sheng Horng. "Wi-Fi SIMO Radar for Deep Learning-Based Sign Language Recognition". In: *IEEE Microwave and Wireless Technology Letters* 34.6 (2024), pp. 825–828. DOI: 10.1109/LMWT.2024.3377712.

[4] J. Landt. "The history of RFID". In: *IEEE Potentials* 24.4 (2005), pp. 8–11. DOI: 10.1109/MP.2005.1549751.

[5] Hong Li et al. "WiFinger: talk to your smart devices with finger-grained gesture". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '16. Heidelberg, Germany: Association for Computing Machinery, 2016, 250–261. ISBN: 9781450344616. DOI: 10.1145/2971648.2971738. URL: https://doi.org/10.1145/2971648.2971738.

[6] Jaime Lien et al. "Soli: ubiquitous gesture sensing with millimeter wave radar". In: *ACM Trans. Graph.* 35.4 (July 2016). ISSN: 0730-0301. DOI: 10.1145/2897824.2925953. URL: https://doi.org/10.1145/2897824.2925953.

[7] Jian Liu et al. "Wireless Sensing for Human Activity: A Survey". In: *IEEE Communications Surveys Tutorials* 22.3 (2020), pp. 1629–1645. DOI: 10.1109/COMST.2019.2934489.

[8] Yongsen Ma et al. "SignFi: Sign Language Recognition Using WiFi". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.1 (Mar. 2018). DOI: 10.1145/3191755. URL: https://doi.org/10.1145/3191755.

[9] Biplov Paneru, Bishwash Paneru, and Khem Narayan Poudyal. "Advancing human-computer interaction: AI-driven translation of American Sign Language to Nepali using convolutional neural networks and text-to-speech conversion application". In: *Systems and Soft Computing* 6 (2024), p. 200165. ISSN: 2772-9419. DOI: 10 . 1016 / j . sasc . 2024 . 200165. URL: https : / / www . sciencedirect . com / science / article / pii / S2772941924000942.

[10] Qifan Pu et al. "Whole-home gesture recognition using wireless signals". In: *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*. MobiCom '13. Miami, Florida, USA: Association for Computing Machinery, 2013, 27–38. ISBN: 9781450319997. DOI: 10.1145/2500423.2500436. URL: https://doi.org/10.1145/2500423.2500436.

[11] Panneer Selvam Santhalingam et al. "Expressive ASL Recognition using Millimeter-wave Wireless Signals". In: *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. 2020, pp. 1–9. DOI: 10 . 1109 / SECON48991 . 2020 . 9158441.

[12] Zengshan Tian et al. "WiCatch: A Wi-Fi Based Hand Gesture Recognition System". In: *IEEE Access* 6 (2018), pp. 16911–16923. DOI: 10.1109/ACCESS.2018.2814575.

[13] Yusen Wang et al. "HearASL: Your Smartphone Can Hear American Sign Language". In: *IEEE Internet of Things Journal* 10.10 (2023), pp. 8839–8852. DOI: 10.1109/JIOT.2022.3232337.

[14] Haotian Zhang et al. "RF-Sign: Position-Independent Sign Language Recognition Using Passive RFID Tags". In: *IEEE Internet of Things Journal* 11.5 (2024), pp. 9056–9071. DOI: 10.1109/JIOT.2023.3322228.

[15] Christian Zieger, Alessio Brutti, and Piergiorgio Svaizer. "Acoustic Based Surveillance System for Intrusion Detection". In: *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 2009, pp. 314–319. DOI: 10.1109/AVSS.2009.49.